International Journal of Advanced Research in Engineering and Technology (IJARET)
Volume 7, Issue 1, Jan-Feb 2016, pp. 105-116, Article ID: IJARET_07_01_013
Available online at
http://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=7&IType=1
Journal Impact Factor (2016): 8.8297 (Calculated by GISI) www.jifactor.com
ISSN Print: 0976-6480 and ISSN Online: 0976-6499

METHODS FOR IMPROVING THE CLASSIFICATION ACCURACY OF BIOMEDICAL SIGNALS BASED ON SPECTRAL FEATURES

Paul Thomas
Professor, Department of Electronics and Communication Engineering,
Mar Baselios College of Engineering and Technology, Thiruvananthapuram

Dr. R.S. Moni
Professor, Department of Electronics and Communication Engineering,
Marian Engineering College, Thiruvananthapuram

ABSTRACT

Biomedical signals are long records of electrical activity within the human body, and they faithfully represent the state of health of a person. Of the many biomedical signals, focus of this work is on Electro-encephalogram (EEG), Electro-cardiogram (ECG) and Electro-myogram (EMG). It is tiresome for physicians to visually examine the long records of biomedical signals to arrive at conclusions. Automated classification of these signals can largely assist the physicians in their diagnostic process. Classifying a biomedical signal is the process of attaching the signal to a disease state or healthy state. Classification Accuracy (CA) depends on the features extracted from the signal and the classification process involved. Certain critical information on the health of a person is usually hidden in the spectral content of the signal. In this paper, effort is made for the improvement in CA when spectral features are included in the classification process. Spectral features are extracted from EEG signal using Multi Wavelet Transform (MWT). Epileptic and Normal cases are classified using k-Nearest Neighbors (k-NN) classifier. Independent Component Analysis (ICA) and Discrete Wavelet Transform (DWT) are used to extract features from ECG signals. These features along with temporal features are used in the classification process. An Artificial Neural Network (ANN) with three hidden layers is used to classify the signal to Ventricular Fibrillation (VF) and non-VF. EMG signal is a train of Motor Unit Action Potentials (MUAP). Dominant MUAP is identified using temporal energy criterion and spectral features are extracted from this using DWT. This method reduces the computational complexity to a large extent. Classification

of the signals in to Amyotrophic Lateral Sclerosis (ALS), Myopathy and Normal is done with k-NN classifier. In all the three cases, CA is found to be better than those based on existing methods. Training data set for classification are selected as those closest to the mean feature vector. This step also contributed to the accuracy of the results.

Key words: Epilepsy, EEG, EMG, MUAP, Myopathy, ALS, ECG, VF, Spectral Features, k-NN, ANN, MWT, DWT, ICA, CA, SVM, EEGLAB, EMGLAB

Cite this Article: Paul Thomas and Dr. R.S. Moni, Methods For Improving The Classification Accuracy of Biomedical Signals Based on Spectral Features. *International Journal of Advanced Research in Engineering and Technology*, **7**(1), 2016, pp. 105-116.
http://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=7&IType=1

1. INTRODUCTION

Among the many biomedical signals, EEG, ECG and EMG alone are considered in this work. EEG is the electrical record of brain's activity observed over a period of 20-40 minutes. These signals are characterized by the firing actions of a large number of neurons in a complex linear system. ECG contains a plethora of information on the normal and pathological physiology of the heart and its health. EMG is the electrical record of the activity of the skeletal muscles during different levels of muscular contractions and they contain a train of MUAPs. It is difficult for physicians to visually examine these long records of biomedical signals to arrive at proper conclusions. During the examination, it is quite likely that they may miss certain features. Also, all the characteristics may not be observable in the time domain signal. Automated analysis and classification of biomedical signals can therefore help the physicians a lot in their diagnostic process.

Epilepsy is the most common Neurological disorder, characterized by the spontaneous occurrence of seizures [1]. Ventricular Fibrillation (VF) is a widely observed cardiac disorder. It is the result of uncoordinated contraction of ventricular muscles, making the ventricles quiver instead of contracting. Amyotrophic Lateral Sclerosis (ALS) is a neuro-degenerative disease that involves the death of neurons. Skeletal muscle weakness is the hallmark of Myopathy. This work is aimed at analyzing and classifying EEG, ECG and EMG signals for these common disorders, with emphasis on spectral features extracted from them.

Various methods are reported for the feature extraction and classification of biomedical signals. Yatindra Kumar etal used Discrete Wavelet Transform (DWT) to extract wavelet entropy values as features from EEG signal and classified these features by t-test statistical method [2]. Subasi also used DWT to extract spectral features from EEG signals [3]. Kalayci and Ozdamar used an Artificial Neural Network (ANN) to classify EEG signals [4]. Zisheng Zhang etal used DWT to extract features from prediction error and used an SVM to classify the features [5]. Ling Guo etal proposed a method for Epileptic seizure detection using Multi Wavelet Transform (MWT) for feature extraction and ANN for classification of EEG signals [6]. Some of the EEG features are Spectral entropy, Spectral squared entropy and Mean spectral magnitude. ANN and SVM are machine learning techniques for classification, where the weights get modified with learning. For ECG signals, the methods for feature extraction and classification include Fourier transform, Discrete cosine transform

(DCT), Hidden Markov model, Fuzzy techniques, ANN, Self organizing map (SOM) and SVM [7]. Different methods are reported for extracting features from EMG signals. They include time domain methods and frequency domain methods. In general, DWT is used to extract spectral features. The two frequency domain approaches for EMG signal analysis are Direct method and MUAP based method. In the Direct method, EMG signal is segmented and each segment is analyzed. In the MUAP based method, the extracted MUAPs are analyzed for information. A lot of information useful for diagnosis can be gathered from MUAPs.

In this work, MWT is used to extract features from EEG signals taken from the dataset described by Andrzejak [1]. The extracted features are separately classified using k-NN, ANN and SVM. The results show that classification of features using k-NN yields a performance better than the other two methods. ICA and DWT are used to extract features from ECG signals taken from MIT database and Physionet. ANN is used for classification. Spectral features of EMG signals are extracted using DWT from dominant MUAP. The extracted features are classified using k-NN and the results are compared with that of SVM. EMG signal available at http://www.emglab.net is used in this work. Block diagram of feature extraction and classification of biomedical signals is shown in fig 1. Selection of training data set is done in a novel way in all the three cases. Data sets closest to the mean of the feature values are used for training. This method is found to improve the CA marginally.

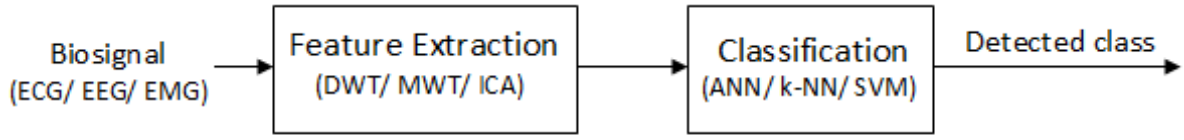

Fig 1. Feature Extraction and Classification

This paper is organized as follows. Materials and methods used in this work are discussed in section 2, the results are discussed in section 3 and the conclusions are made in section 4.

2. MATERIALS AND METHODS

2.1. Sources of the Biomedical signals

Biomedical signals taken from online data bases are used in this work. The EEG dataset described by Andrzejak etal [1] is used for the classification of Epilepsy and Normal cases. This database contains five sets of data (S, F, N, O, Z), recorded using 10-20 electrode placement scheme. Each set contains 100 EEG signals; each signal has 4096 samples corresponding to a duration of 23.6 seconds. Z contains EEG records of five healthy subjects who were awake, relaxed and eyes open. O contains EEG records of five healthy subjects who were awake, relaxed and eyes closed. F contains EEG records taken from epileptogenic zone in patients with Epilepsy. N contains EEG records taken from hippocampal formation in such patients. S contains EEG records taken during seizure. Z and F data sets are used in this work. Samples of EEG signal from sets Z and F are plotted in Fig 2.

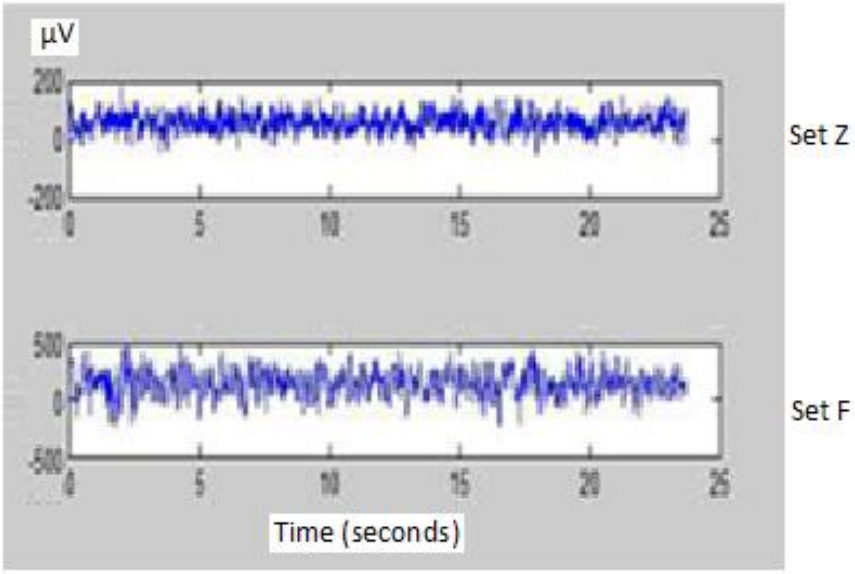

Fig 2. Samples of EEG signals

67 ECG records from MIT data base and the Physionet repository (30 VF and 37 non-VF) are used in this work. Part of this data set is used for training the ANN. The ECG signal is pre- processed using Moving average filter. A typical ECG signal from the repository is plotted fig 3.

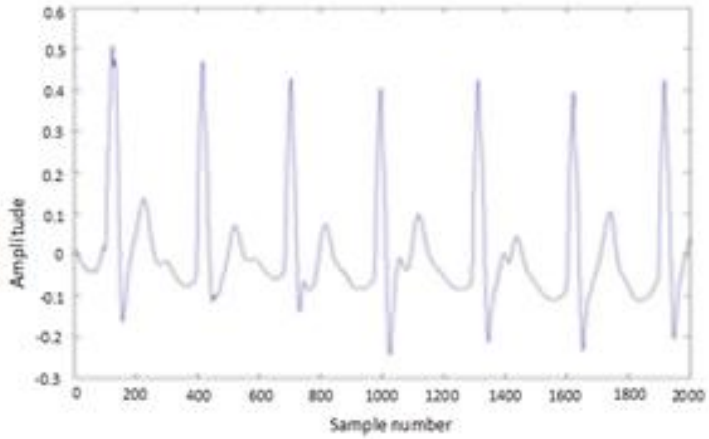

Fig 3. Sample ECG signal from MIT database

EMG signals are taken from http://www.emglab.net, which contains clinical signals recorded by Nikolic. It contains three data sets which are combinations of Normal subjects, ALS patients and Myopathy patients. They are Myo-normal dataset, ALS-normal dataset and 3-class dataset respectively. 3-class dataset consists of 50 myopathy signals of 7 patients, 50 ALS signals of 7 patients and 150 normal signals of 10 subjects. The cut off frequencies of the high pass and low pass filters used are set at 2 Hz and 10 kHz respectively. Each signal is recorded in binary format at 23,438 samples/sec for 11.2 seconds duration. Sample EMG signals of ALS, Myopathy and Normal are plotted in fig 4.

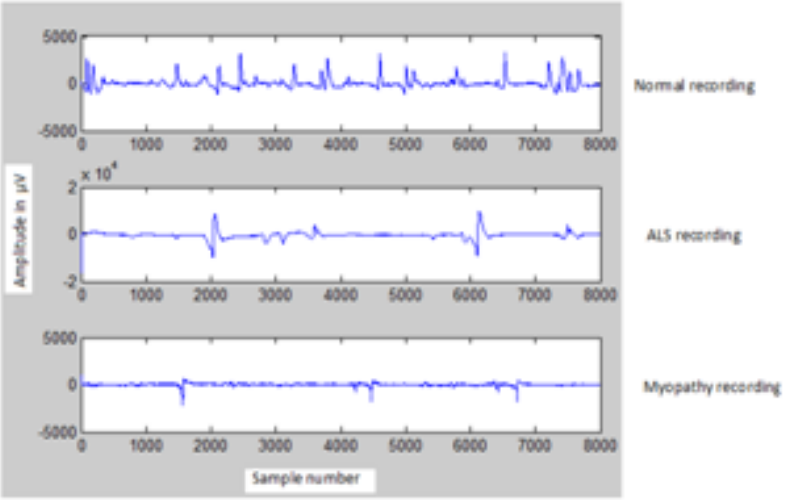

Fig 4. Samples of EMG signals

2.2. Tools used in the work

MATLAB R2013a is the development platform used in this work. It is an interactive environment that helps to analyze data, create models and applications. The tool boxes of neural network, SVM, wavelet transform, ICA and signal processing are used. The open source software packages EEGLAB, EMGLAB and Biosig are integrated into the MATLAB environment. EMGLAB is used to decompose EMG signals into MUAPs. It supports automatic decomposition, manual editing and verification of results. EEGLAB is used for processing EEG data. It supports independent component analysis (ICA), artifact rejection, and several modes of data visualization. EEGLAB allows import of EEG data in many different file formats. BioSig helps in data acquisition, artifact processing, feature extraction, classification, modeling and visualization of the biomedical signal.

2.3. Features and methods of feature extraction

2.3.1. EEG Features

EEG signals from sets F and Z (200 EEG blocks of 23.1 s each) are used. Number of samples in each block is 4096. Spectral features are extracted from EEG signal using Gernoimo-Hardin Massopust (GHM) multi-wavelet [6]. GHM has two wavelet functions and two scaling functions. Signal analysis produces two sub bands in the high frequency and two sub bands in the low frequency. The spectral features are Spectral entropy (P_{se}), Spectral squared entropy (P_{sse}) and Mean spectral amplitude (M_{ave}).

$$M_{ave} = \frac{1}{N}\sum_k |X_k|; \qquad X_k \text{ are the approximat e coefficien ts}$$

$$P_{se} = -\sum_k q_k \log q_k, \qquad where \;\; q_k = \frac{|X_k|}{\sum_k |X_k|}$$

$$P_{sse} = -\sum_k r_k \log r_k, \qquad where \;\; r_k = \frac{|X_k|^2}{\sum_k |X_k|^2}$$

2.3.2. ECG features

2.3.2.1. The ECG temporal features are extracted from a signal segment of 3 seconds duration [7]. The features are (1) Threshold crossing interval (TCI), the average time duration between consecutive threshold crossings (2) Threshold crossing count (TCC), the number of samples crossing a threshold (3) Exponent Crossing Count (ECC), the number of crossings segment with an exponent drawn from the ECG sample with maximum amplitude and (4) Mean absolute value (MAV) of the signal.

2.3.2.2. ECG spectral features are (1) Power spectral density (PSD), spectral amplitudes in sub bands extracted with Daub8 wavelet decomposition (2) Median frequency (MF) of the power spectrum of the ECG signal and (3) Sample entropy (SE), the negative natural logarithm of the conditional probability that two sequences similar for 'm' points remain similar at the next point (A lower value indicates more self-similarity) [8].

2.3.2.3. ICA features are derived from the ICA coefficients obtained when the ECG signal is transformed in to linearly independent basis vectors [9].

2.3.3. EMG Features

EMG signal contains a train of MUAPs. The signal is decomposed into MUAPs using EMGLAB. Energy content of MUAPs are low in Myopathy patients and high in ALS patients, compared to Normal. The dominant MUAP is identified as that with the highest temporal energy and spectral features are extracted from this.

2.3.3.1. Direct extraction of spectral features

Periodogram is used to compute the power spectrum of the dominant MUAP. Mean power, Total power, First, Second and Third spectral moments, Peak power, Median frequency and Mean frequency are the features normally considered. CA of ALS and Myopathy are found to be low with these eight features. Further studies revealed that CA is improved with four out of the eight features, namely Peak power, Total power, Mean power and Median frequency.

2.3.3.2. DWT based extraction of spectral features

Dominant MUAP is one stage decomposed using Daub-2 wavelet [10]. The spectral features are extracted from the approximate coefficients. It is observed that spectral features extracted from detail coefficients reduces the CA and that higher levels of decomposition does not lead to any improvement in CA.

2.4. Methods of feature classification

2.4.1. Classifiers

A k-NN classifier computes a distance function between the features belonging to the input signal and k neighboring patterns of the training data [11]. In this method, Euclidean distance is computed to find out the class to which the input pattern is closest. A large value for k can give improved performance, but drastically increase the computational complexity. Hence, k is chosen as 5. ANN classifier consists of heavily interconnected neurons, the processing elements. Multi-layer perceptron (MLP) with back propagation algorithm is shown in fig 5. A three-layer MLP is used in this work. Training of ANN requires several iterations. The connection weights are initially chosen at random. For each of the training data, the error is computed in a forward pass. The weights are then updated based on Gradient-descent algorithm.

SVM is a binary classifier that minimizes the empirical classification error and maximizes the geometric margin. They build a boundary that separate data using a linear hyper plane. The online data base contains pre-classified signals belonging to the normal and the diseased conditions and they are used for training. The criterion of defining the hyperplane for this classification is maximum Euclidian distance from the plane.

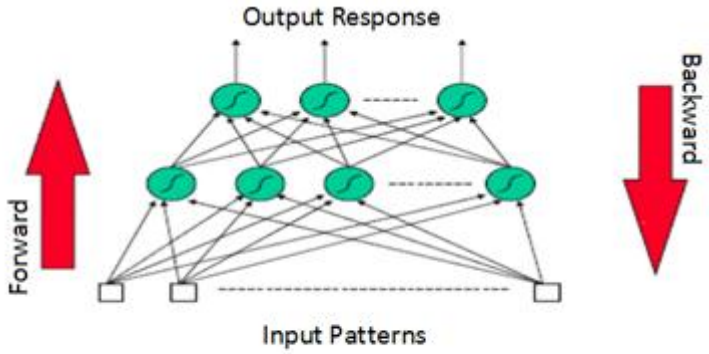

Fig 5. Multilayer perceptron

2.4.2. Selection of training data

It is observed that the CA is marginally better with a structured selection of training data, instead of a random pick up. The approach here is to reduce the number of outliers in the training data.

Following method is proposed for selecting the training data: (1) Compute the mean of all feature vectors and the Euclidean distance of each feature vector from this mean [12]. (2) Arrange the feature vectors in the ascending order of this distance from the mean vector. (3) Select the feature vectors closest to the mean vector for training purpose.

2.4.3. EEG feature classification

The extracted features are classified for the normal and the epileptic seizure cases. The classification is done separately using k-NN, ANN and SVM and is observed that the performance with k-NN is the best. Block diagram of the method is shown in Fig 6.

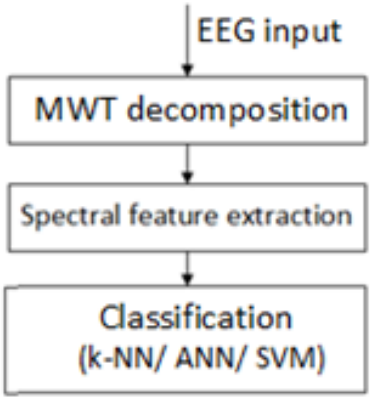

Fig 6. EEG classifiaction

2.4.4. ECG feature classification

ANN is used to classify the ECG features into VF and non-VF categories. Classification is attempted with three different feature sets. The first feature set has spectral and temporal features. The second feature set has ICA features and the third one has ICA, spectral and temporal features. Block diagram of the classification scheme is shown in fig 7.

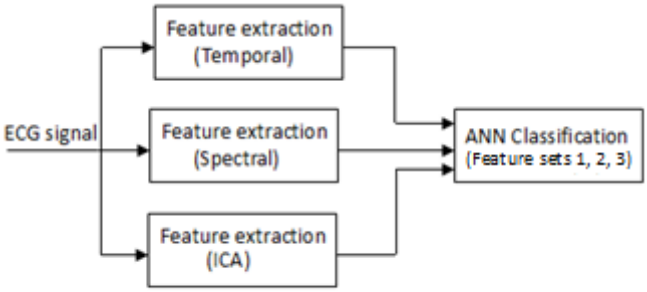

Fig 7. ECG classification

2.4.5. EMG feature classification

In the general method of classification based on MUAPs, all MUAPs have the same level of importance. The method has the limitation that the number of extracted MUAPs and their characteristics vary significantly from one EMG signal to another. In this work, it is proposed to use the MUAP with highest energy content (dominant MUAP) for analysis [12].

Energy of an MUAP with N samples: $e_f = \sum_{n=0}^{N-1} |x(n)|^2$

Here, feature extraction is done only from the dominant MUAP. The computational complexity is much reduced, since we are extracting features from only one MUAP. Since SVM is a binary classifier, it can work on only two classes of signals. In order to have a comparison of the CAs achieved with SVM and k-NN classifiers, k-NN is operated on two classes of signals at a time. Block diagram of EMG classification scheme is shown in the fig 8.

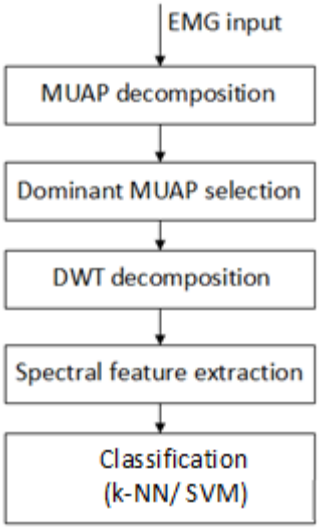

Fig 8. EMG classification

2.4.6. Performance evaluation of classification

The statistical parameters used for performance evaluation are Sensitivity (SE), Specificity (SP) and CA. The disease conditions considered are Epilepsy with EEG signal, ALS and Myopathy with EMG signal and VF with ECG signal. SE is the ratio of correct detection of disease condition. SP is the ratio of correct detection of non-disease condition. CA is the ratio of correct classification of disease and non-disease conditions.

$$SE = \frac{TP}{TP + FN}, \qquad SP = \frac{TN}{TN + FP}$$

$$CA = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP is the number of correctly classified disease cases, FN is the number of wrongly classified disease cases, TN is the number of correctly classified Normal cases and FP is the number of wrongly classified Normal cases

3. RESULTS AND DISCUSSIONS

The values of SP, SE and CA are computed by comparing the results obtained in classification with diagnostic information available. Classification for the epileptic seizure/ the normal is carried out with EEG signals taken from the dataset described by Andrzejak. EEG signals from sets Z and F (200 EEG blocks) are used in this work. Feature extraction is carried out with GHM multi wavelet. Classification is carried out with k-NN and the results are found to be better than that obtained using ANN and SVM. It is seen that k-NN provides the highest CA. The result is tabulated in TABLE 1.

TABLE 1 Performance of EEG classification

Classifier	SP (%)	SE (%)	CA (%)
SVM	96	88	92
ANN	95	97	96
k-NN	97	96	96.5

The performance obtained with Multi wavelets is compared with that of Scalar wavelets in the k-NN classification environment. Multi-wavelets GHM, Chui-Lian (CL) and the scalar wavelets Daub2 and Daub8 are separately used for feature extraction. The accuracy obtained with the different wavelets is shown in fig 8. Results show that CA is better when multi-wavelets are used for feature extraction, compared to scalar wavelets.

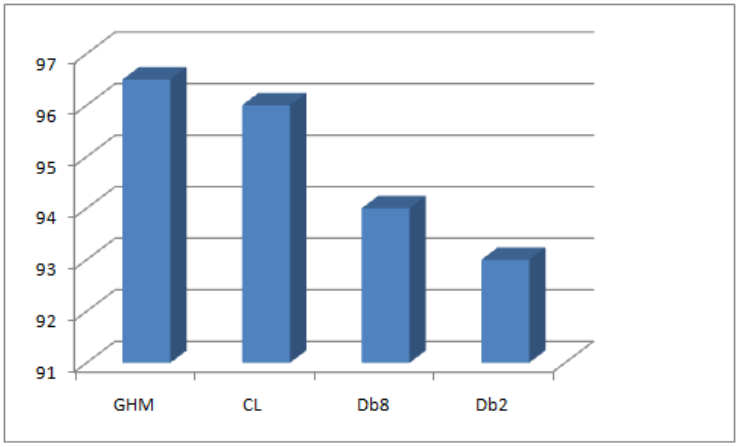

Fig 8. CA of EEG signal with different wavelets

Classification for VF/ non-VF is carried out with ECG signals taken from MIT database and Physionet. Temporal, spectral and ICA features extracted are classified using ANN classifier. Classification is attempted on three sets of features. Spectral and temporal features constitute the first set, ICA features constitute the second set and all the features combined constitute the third set. Performance of ECG classification is tabulated in TABLE 2. It can be inferred that the SE, SP and CA are best with Method 3.

TABLE 2 Performance of ECG classification

	SE (%)	**SP (%)**	**CA (%)**
Method 1	96.7	83.8	89.6
Method 2	76.7	97.3	88
Method 3	96.7	97.3	97

Classification for ALS/ Myopathy/ Normal is carried out on 3-class EMG dataset. This dataset consists of 50 ALS signals, 50 Myopathy signals and 150 Normal EMG signals of 11.2 seconds duration each. Feature extraction from dominant MUAP is carried out in two different ways (1) using Periodogram and (2) using Daub2 wavelet. Classification is done with SVM and k-NN classifier separately. The CA with k-NN classifier is better than that with SVM. The performance of EMG classification with k-NN classifier is tabulated in TABLE 3. The results for two different methods of feature extraction are listed. The results show that CA is better with DWT used for feature extraction, compared to periodogram method of feature extraction.

TABLE 3 Performance of EMG classification

	SP (%)	**SE of Myo (%)**	**SE of ALS (%)**	**CA (%)**
Periodogram method	88	78	92	86.8
DWT method	92	76	94	89.2

4. CONCLUSION

EEG feature extraction with GHM multi wavelet and classification using k-NN provides classification accuracy better than those with other methods. ECG feature extraction using ICA and DWT and classification using ANN provides good CA for Ventricular Fibrillation. EMG signal classification scheme with spectral features extracted from dominant MUAP using DWT, provides good CA. The computational complexity in feature extraction is also less since we concentrate on the dominant MUAP alone. A new method of selecting the training data set improves the classification accuracy in all the cases.

More studies can be carried out on the use of MWT and wavelet packets for feature extraction. Studies can also be carried out on classifying more disease conditions for these biomedical signals. Further investigation can be carried out on ANN based classification schemes to understand why the performance obtained with k-NN classifier is better for EEG and EMG. It can also be explored if CA can be improved by assigning non-uniform weights to the various spectral features. Possibility of developing a generalized scheme for ECG, EEG and EMG signal analysis can also be explored.

REFERENCES

[1] Andrzejak R.G, Lehnertz K, Mormann F, Rieke C, David P, Elger C.E Indications of nonlinear deterministic and finitedimensional structures in time series of brain electrical activity: dependence on recording region and brain state, Physical Review E, 2001; vol. 64(6):061907-1–8.

[2] Yatindra Kumar, M.L Dewal and R.S Anand, Wavelet Entropy Based EEG Analysis for Seizure Detection, Biomedical Engineering and Informatics (BMEI), 2013 International Conference pp. 878-882.

[3] Subasi A, EEG signal classification using wavelet feature extraction and a mixture of expert model, Expert Systems with Applications 2007; 32(4): 1084–93.

[4] Kalayci T, Ozdamar O. Wavelet preprocessing for automated neural network detection of EEG spikes, IEEE Engineering in Medicine and Biology Magazine 1995; 14(2):160-6.

[5] Zisheng Zhang, Student Member, IEEE, and Keshab K. Parhi, Fellow, IEEE, Seizure Detection using Wavelet Decomposition of the Prediction Error Signal from a Single Channel of Intra-Cranial EEG, 2014, EMBC, 36th Annual International Conference of the IEEE, pp. 4443-4446, , 2014.

[6] Ling Guo, Daniel Rivero, Alejandro Pazos. Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks; Journal of Neuroscience Methods 193 (2010) 156–163.

[7] Felipe Alonso-Atienza, Eduardo Morgado, Lorena Fern´andez-Mart´ınez, Arcadi Garc´ıa-Alberola and Jose´ Luis Rojo-A´lvarez, "Detection of Life-Threatening Arrhythmias Using Feature Selection and Support Vector Machines, IEEE Trans. Biomedical Engineering, vol.61, no. 3, pp. 832-840, Mar. 2014.

[8] A. Amann, R. Tratnig, and K. Unterkofler, A new ventricular fibrillation detection algorithm for automated external defibrillators, Proc Computers in Cardiology, pp. 559–562, Sep. 2005.

[9] A. Hyvarinen and E. Oja, A fast fixed-point algorithm for independent component analysis, Neural Computation, vol. 9, no. 7, pp. 1483–1492, 1997.

[10] Angkoon Phinyomark , Pornchai Phukpattaranont, Chusak Limsakul, Feature reduction and selection for EMG signal classification, Expert systems with applications 39 (2012) 7420-7431.

[11] Oliver Sutton, Introduction to k Nearest Neighbour Classification and Condensed Nearest Neighbor Data Reduction, February, 2012.

[12] Anu V.S, Paul Thomas, Superior performance by k-NN based method in classifying Epileptic EEG signals, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol.4, Issue 6, pp 2320-3765, June 2015